Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP2003/013394

International filing date: 28 November 2003 (28.11.2003)

Document type: Certified copy of priority document

Document details: Country/Office: EP
Number: 02026770.4
Filing date: 02 December 2002 (02.12.2002)

Date of receipt at the International Bureau: 30 November 2006 (30.11.2006)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

World Intellectual Property Organization (WIPO) - Geneva, Switzerland
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse

Europäisches Patentamt	European Patent Office	Office européen des brevets

EP/03/13394

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

02026770.4

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02026770.4
Demande no:

Anmeldetag:
Date of filing: 02.12.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

DEUTSCHE THOMSON-BRANDT GMBH
Hermann-Schwer-Strasse 3
78048 Villingen-Schwenningen
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Method for describing the composition of audio signals

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G10L/

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SI SK

PD020112-Ri-021202

Method for describing the composition of audio signals

The invention relates to a method for describing the composition of audio signals, especially for the spatialization of MPEG-4 encoded audio signals in a 3D domain.

The MPEG-4 Audio standard as defined in ISO/IEC 14496-3 and 14496-1 facilitates a wide variety of applications by supporting the representation of audio objects. For the combination of the audio objects additional information - the so-called scene description - determines the placement in space and time and is transmitted together with the coded audio objects.

A scene description is structured hierarchically and can be represented as a graph, wherein leaf-nodes of the graph form the separate objects and the other nodes describes the processing, e.g. positioning, scaling, effects. A node named "Sound" allows spatialization of the audio signal in a 3D domain. A further node with the name "Sound2D" only allows spatialization on a 2D screen. The use of the "Sound" node in a 2D graphical player is not specified due to different implementations of the properties in a 2D and 3D player.

The Sound2D node is defined as followed:

```
Sound2D {
    exposedField    SFFloat   intensity  1.0
    exposedField    SFVec2f   location   0,0
    exposedField    SFNode    source     NULL
    field     SFBool    spatialize  TRUE
}
```

and the Sound node is defined as followed:

```
Sound {
    exposedField        SFVec3f   direction     0, 0, 1
    exposedField        SFFloat   intensity     1.0
    exposedField        SFVec3f   location  0, 0, 0
    exposedField        SFFloat   maxBack   10.0
    exposedField        SFFloat   maxFront  10.0
    exposedField        SFFloat   minBack   1.0
    exposedField        SFFloat   minFront  1.0
    exposedField        SFFloat   priority  0.0
    exposedField        SFNode    source    NULL
    field       SFBool    spatialize  TRUE
}
```

In the following the general term for all sound nodes (Sound2D, Sound and DirectiveSound) will be written in lower-case e.g. 'sound nodes'.

Please pay attention to the fact that the Sound node is actually a *3D* node.

In the simplest case the Sound or Sound2D node is connected via an AudioSource node to the decoder output. The sound nodes contain the *intensity* and the *location* information. From the audio point of view a sound node is the final node before the loudspeaker mapping. In the case of several sound nodes, the output will be summed up. From the systems point of view the sound nodes can be seen as an entry point for the audio sub graph. A sound node can be grouped with non-audio nodes into a Transform node that will set its original location.

With the *phaseGroup* field of the AudioSource node, it is possible to mark channels that contain important phase relations, like in the case of "stereo pair", "multichannel" etc. A mixed operation of phase related channels and non-

phase related channels is allowed. A *spatialize* field in the sound nodes specifies whether the sound shall be spatialized or not. This is only true for channels, which are not member of a phase group.

The Sound2D can spatialize the sound on the 2D screen. The standard said that the sound should be spatialized on scene of size 2m x 1.5m in a distance of one meter. This explanation seems to be ineffective because the value of the location field is not restricted and therefore the sound can also be positioned outside the screen size. We suppose to remove the confusing description.

The Sound and DirectiveSound node can set the *location* everywhere in the 3D space. The mapping to the existing loudspeaker placement can be done using simple amplitude panning or more sophisticated techniques.

Both Sound and Sound2D can handle multichannel inputs and basically have the same functionalities, but the Sound2D node cannot *spatialize* a sound other than to the front.

From games, cinema and TV applications we know, that it makes sense to provide the end user with a fully spatialized "3D-Sound" presentation, even if the video presentation is limited to a small flat screen in front.

A first proposal is to add Sound and Sound2D to all scene graph profiles, i.e. add the Sound node to the SF2DNode group.

But, one reason for not including the "3D" sound nodes into the 2D scene graph profiles is, that a typical 2D player is not capable to handle 3D vectors (SFVec3f type), as it would be required for the Sound *direction* and *location* field.

Another reason is that the Sound node is specially designed for virtual reality scenes with moving listening points and attenuation attributes for far distance sound objects. For this the Listening point node and the Sound *maxBack*, *maxFront*, *minBack* and *minFront* fields are defined.

The proposal is to extend the old Sound2D node or to define a new Sound2Ddepth node. The Sound2Ddepth node should be similar the Sound2D node but with an additional *depth* field.

```
Sound2Ddepth {
    exposedField        SFFloat intensity   1.0
    exposedField        SFVec2f location0,0
    exposedField        SFFloat depth    0.0
    exposedField        SFNode  source   NULL
    field       SFBool  spatialize  TRUE
}
```

The *intensity* field adjusts the loudness of the sound. Its value ranges from 0.0 to 1.0, and this value specifies a factor that is used during the playback of the sound

The *location* field specifies the location of the sound in the 2D scene.

The *depth* field specifies the depth of the sound in the 2D scene using the same coordinate system than the location field. The default value is 0.0 and it refers to the screen position.

The *spatialize* field specifies whether the sound shall be spatialized. If this flag is set, the sound shall be spatialized with the maximum sophistication possible.

The same rules for multichannel audio spatialization apply to the Sound2Ddepth node as to the Sound (3D) node.

Using the Sound2D node in a 2D scene allows presenting surround sound, as the author recorded it. It is not possible to *spatialize* a sound other than to the front. Spatialize means moving the location of a monophonic signal due to user interactivities or scene updates.

With the new proposal of a Sound2Ddepth node it is possible to *spatialize* a sound also in the back, at the side or above of the listener. Supposing the audio presentation system has the capability to present it.

The invention allows the spatialization of the audio signal in a 3D domain, even if the player is restricted to 2D graphics.

A further embodiment is as follows:

The audio nodes 'Sound', 'Sound2D' and 'directiveSound' described in the MPEG-4 audio standard have besides the functionality to present phase-related multichannel signals (>1channel), the functionality to present accumulated monophonic sounds everywhere in the 2D or 3D space. Their position in these spaces can be addressed by using the 'location'-field of these nodes. The location is a 2D- or 3D-vector. If the 'spatialize'-field of these nodes is set to 'true', the sound will be spatialized depending on the 'location'. For the spatialization process different algorithms can be used, for example the amplitude panning.
With these techniques it is only possible to position a sound as a point sound source. The sound has no 'width'.

Different methods are actually discussed to position a sound in a virtual reality world space, with a noticeable width. The approach of these algorithms is to describe a size or a

shape in a 3D coordinate system.
There has been no approach for a user-oriented system. The following solution is to insert two new fields into the Sound and Sound2D nodes to control the width of the sound with an opening-angle relative to the listener: The angle has a vertical and a horizontal component, 'widthHorizontal' and 'widthVertical', ranging from 0...2π with the location as its center. The widthHorizontal is shown in Fig. 3. The widthVertical is similar to this with a 90 degree x-y-rotated relation.

The width of a sound can be generated for example as described in DE - A - 196 32 734 'methods and apparatus for generating a multichannel signal from a mono signal'.

An application is for example the monophonic transmission of the violins in an orchestra instead its transmission as stereo signals, coupled to a fixed loudspeaker layout and no possibility to position it at a desired location.

Claims

1. Method for describing the composition of audio signals, which are encoded as separate audio objects, wherein the arrangement and the processing of the audio objects in a sound scene is described by nodes arranged hierarchically in a scene description, wherein a node allows spatialization on a 2D screen using a 2D vector, **characterized by** describing a 3D position of an audio object using said 2D vector and a 1D value describing the depth of said audio object.

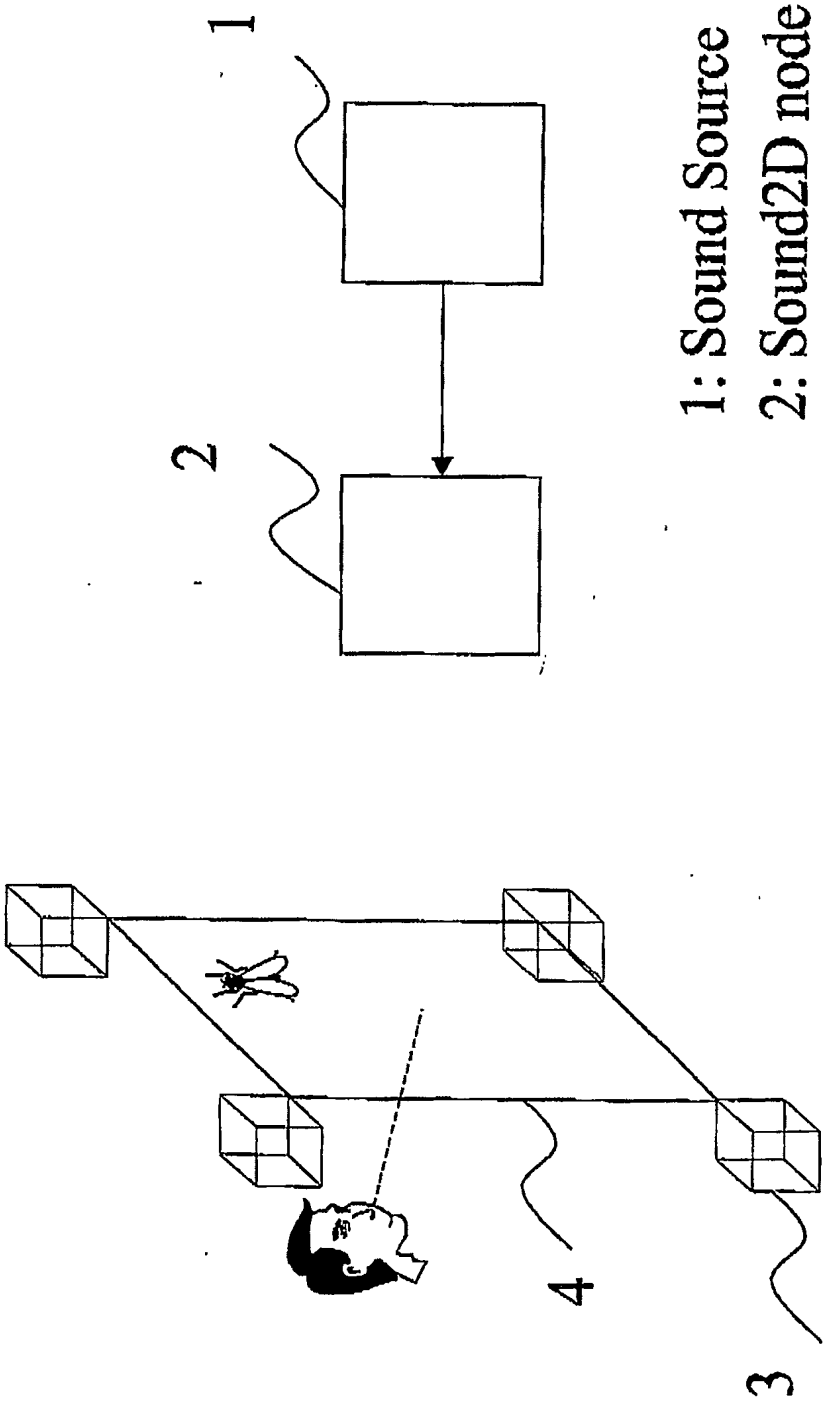
2. Method according to claim 1, **characterized by** controlling the width of an audio object by using two additional fields inserted into the Sound and Sound2D nodes describing an opening-angle relative to the listener.

Abstract

Method for describing the composition of audio signals, which are encoded as separate audio objects. The arrangement and the processing of the audio objects in a sound scene is described by nodes arranged hierarchically in a scene description. A node specified only for spatialization on a 2D screen using a 2D vector describes a 3D position of an audio object using said 2D vector and a 1D value describing the depth of said audio object.

In a further embodiment two new fields are inserted into the Sound and Sound2D nodes to control the width of the sound with an opening-angle relative to the listener.

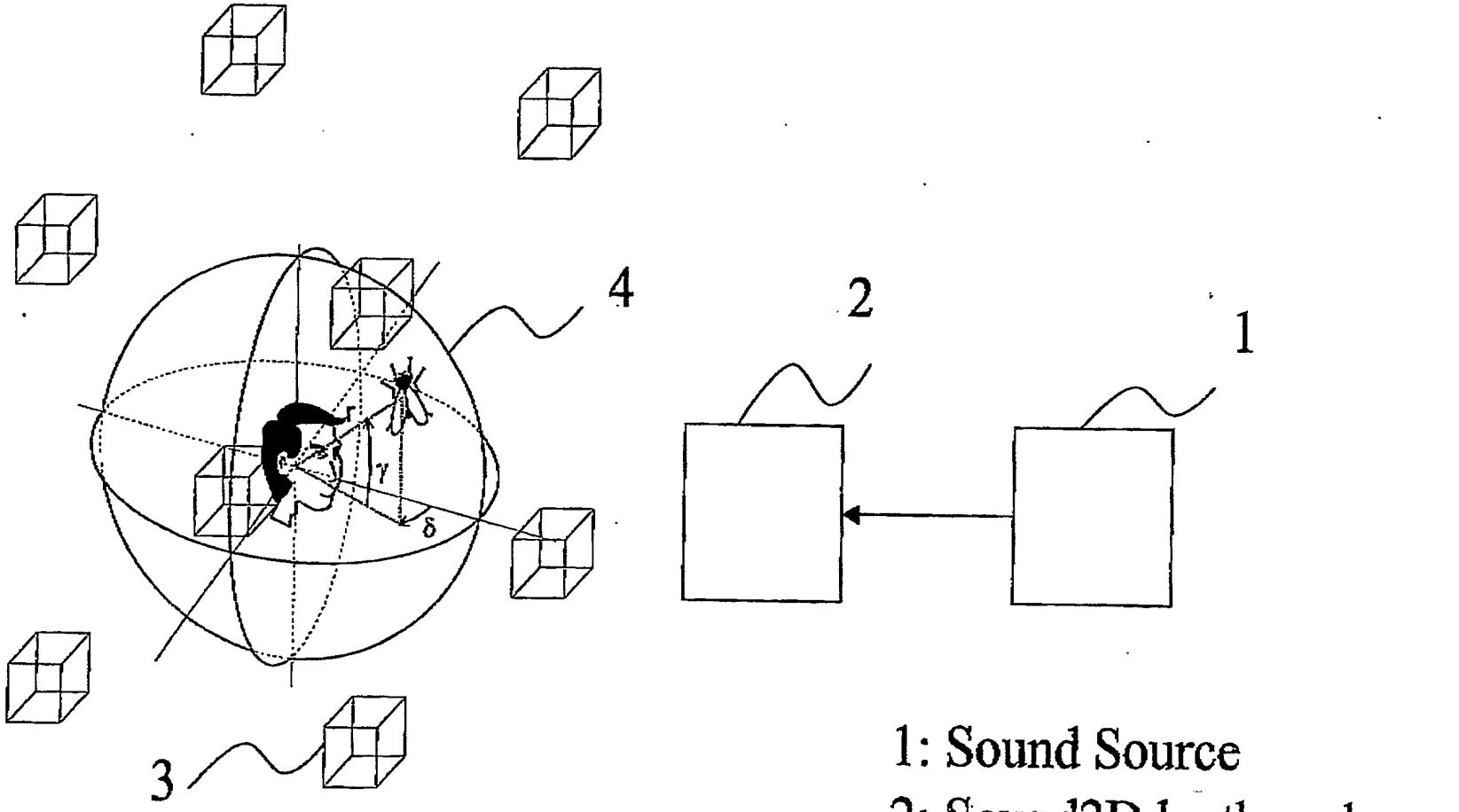
Fig. 1



Empfangszeit 2.Dez. 13:22

2.DEZ.2002 13:20

Fig. 2



1: Sound Source
2: Sound2Ddepth node
3: Speaker
4: 3D Sound Reproduction

-- NR.783 . S.16/17 -

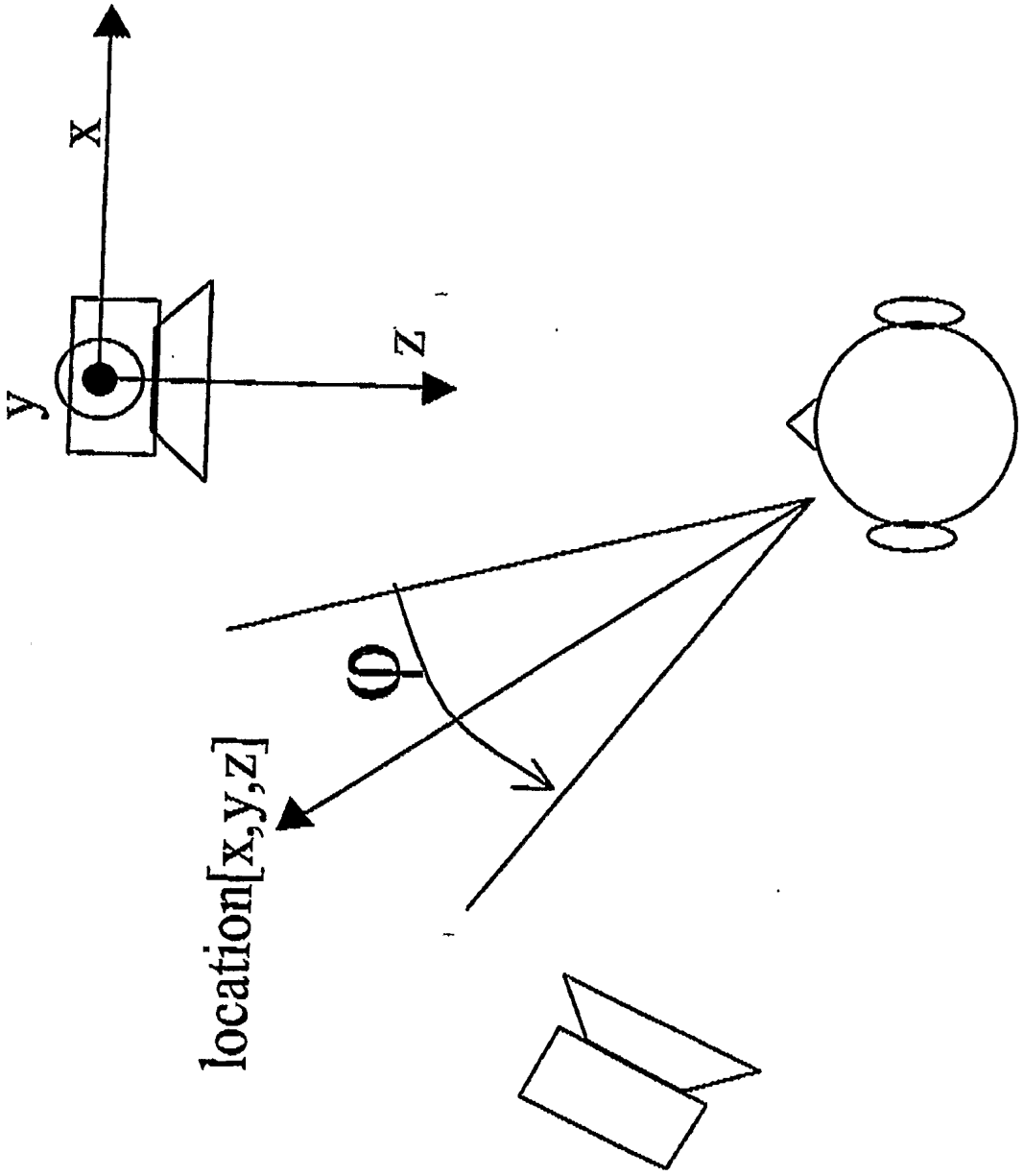


Fig. 3: Definition of the 'widthHorizonal' with the location as its center.